View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UMP Institutional Repository

DETECTION OF PRESENCE MARKS IN STEGOCONTENT

SIA CHIA MENG CA10008

THESIS SUBMITTED IN FULFILMENT OF THE DEGREE OF COMPUTER SCIENCE

FACULTY OF COMPUTER SYSTEM AND SOFTWARE ENGINEERING

2013

ABSTRACT

Steganography is an art of hiding message by embedding message within other types of media for example text, audio or image files. The main purpose of Steganography is attempts to hide the existence of communication. For steganography, the users totally do not want to let anyone know they are sending messages. The objectives of this project are study current steganography tools and the methods used for detection of presence marks in stegocontent, develop steganalysis tools which implement with graphical user interface (GUI), and verify and detect the marks in stegocontent in image files to assist investigation process. Peak Signal-to-Noise Ratio (PSNR) steganalysis is chose as the method used in this application software because it is use for detect various Least Significant Bits (LSB) modification techniques. LSB is the method that most commonly used in steganography to hide message.

ABSTRAK

Steganografi adalah seni menyembunyikan mesej dengan menerapkan mesej di dalam lain-lain jenis media contohnya teks, audio atau imej. Tujuan utama steganografi adalah untuk menyembunyikan kewujudan komunikasi. Bagi steganografi, pengguna benar-benar tidak menginginkan orang lain tahu bahawa mereka menghantar mesej. Objektif projek ini adalah mengaji alat-alat steganografi semasa dan kaedah yang digunakan untuk mengesan tanda-tanda kehadiran di stegocontent, membangunkan alat steganalysis yang melaksanakan dengan Graphical User Interface (GUI), dan mengesahkan atau mengesan tanda-tanda dalam stegocontent dalam fail imej untuk membantu proses penyiasatan . Peak Signal-to-Noise Ratio (PSNR) steganalysis adalah memilih sebagai kaedah yang digunakan dalam aplikasi perisian ini kerana ia digunakan untuk mengesan pelbagai Least Significant Bits (LSB) teknik pengubahsuaian. LSB adalah kaedah yang paling biasa digunakan dalam steganografi untuk menyembunyikan mesej.

TABLE OF CONTENT

LIST OF TABLE

LIST OF FIGURES

LIST OF ABBREVIATIONS

BMP	Bitmap format
DCT	Discrete Cosine Transform
DFD	Data Flow Diagram
DWT	Discrete Wavelet Transform
GIF	Graphics Interchange Format
GUI	Graphical User Interface
JPEG	Joint Photographic Experts Group
LSB	Least Significant Bits
MSE	Mean Square Error
PNG	Portable Network Graphics
PoV	Pairs of Value
PSNR	Peak Signal-to-Noise Ratio
QIM	Quantization Index Modulation
RQP	Raw Quick Pairs
RS	Regular and Singular groups

CHAPTER 1:

INTRODUCTION

1.1 Introduction

Steganography is an art of hiding message by embedding message within other types of media for example text, audio or image files. According to Kumar & Pojar (2010), steganography is not same as cryptography. Steganography is about secrecy and cryptography is concern on privacy. The cryptography is used to protect the information from reveal to public. For steganography, the users totally do not want to let anyone know they are sending messages. Few centuries ago, the first uses of steganography by Demaratus, the king of Sparta to send a warning about a forthcoming attack to Greece by writing the message on the wood and cover the message with wax upon to hide the secret message. There is also the ancient people hiding the message by tattooed the message or map on shaved head of a person. The message is then become hidden when the hair is grown and they will shaved the head again when they want to see the message. Those are some of the techniques or method that ancient people use to hiding the message. Now, steganography become more popular due to internet and multimedia as the internet makes the message transfer fast and free. Nowadays, steganography can be used for water marking where a message is hidden in the "carrier" so that its sources can be track or verify (Curran & Devitt, 2008). However, some of the users use steganography technique to commit crime because it is hard to detect. There is some of the steganalytic software used to detect the media that have hidden message but it is not fully implement with GUI. This makes new users hard to learn to get familiar to the functions and operations of the software as they have insufficient technical knowledge. Besides that, there are limited numbers of steganalytic software in the market. In conclusion, although the use steganography technique is not intend to doing something bad, there are cases that show that some people misuse it to commit crime. So, this steganalytic software can help the investigation on stegocontent to prevent crime. It also make the user can use it easily and convenient.

1.2 Problem Statement

The basic structure of Steganography is made up of three components: the “carrier”, the message, and the key. The main purpose of Steganography is attempts to hide the existence of communication. According to Cole (2003), three principles can be used to measure the effectiveness of a steganography technique. Those principles are amount of data, difficultly of detection and difficulty of removal. For amount of data, explanation will be the more data can be hided, more great the technique. Difficulty of detection relates to how easy it is for someone to detect that a message has been hidden. Once the amount of data hidden increases in a file, the risk the message be detected also becomes higher. Difficulty of removal suggests that the hidden data cannot be able to remove easily. Actually steganography can be used to protect intellectual property from digital robbery (copyrights) by injecting the copyright marks and serial number in electronic medium such as books and audio. However, there are some irresponsible users abused steganography techniques to commit crimes. For example, criminals can communicate with each other secretly by hiding the messages using steganography techniques to commit crimes like selling and transfer drugs and avoid getting caught by polices. It is very difficult to investigate those crimes since there are too many techniques to hide the message. Besides that, there are limited numbers of software or program that help to detect the marks of stegocontent. Furthermore, most of the steganography analysis software are lack of Graphical User Interface (GUI) so many new users does not understand how to operate them and it is complex to use.

1.3 Objectives

This project comprehends the following objectives:

i. To study current steganography tools and the methods used for detection of presence marks in stegocontent.
ii. To develop steganalysis tools which implement with graphical user interface (GUI).
iii. To verify and detect the marks in stegocontent in image files to assist investigation process.

1.4 Scope

i. This steganalysis application software targets all computer users.
ii. All of the computer users can use this tool to detect stegocontent for precaution or security purposes.
iii. This steganalysis tool is focus on the detection of stegocontent in image files (JPEG/JPG, PNG) only.
iv. This steganalysis tool detects only Least Significant Bits (LSB) modification in image files.

1.5 Thesis Organization

This project report includes chapter 1, chapter 2, and chapter 3. Chapter 1 provides a broad overview of the thesis. In chapter 1, it includes introduction, problem statements, objectives, scopes and thesis organization. Chapter 2 reviews the previous research of steganography and comparison of steganography and steganalysis methods. Chapter 3 describes the approach used to develop the application. It includes prototype interface, use case diagram, flow chart of the application software.

Chapter 2:

Literature Review

In recent years the growth in the quantity of available Steganography tools on Internet can be seen obviously. In short, Steganography is the art or technique to hide messages within types of media. The goal of steganography is to keep the existence of a message hidden or to hide the fact that communication is taking place. In contrast, the goal of cryptography is to make the messages cannot be understood (Druid, 2006). Nowadays, Internet is widely use to transfer and store information or messages. This means that Internet can be considered to be a storehouse of steganographic materials (Callinan & Kemick, 2006). It is believed that criminals and terrorist organizations may be communicating secretly through the use of steganography. Therefore, Steganalysis, the techniques of detecting hidden messages using Steganography is necessary to detect hidden data. However, some steganographic techniques are particularly difficult to detect without the original sources (Johnson & Jajodia, 1998).

2.1 Introduction of Image Steganography

According to Curran & Devitt(2008), digital images are the most widely used medium for steganography today and it take advantage of our limited visual perception of colour. The most popular image formats on the internet are the graphics interchange format (GIF), joint photographic experts group (JPEG) format, the portable network graphics (PNG) format and the bitmap format (BMP). Queirolo(2006) states that large images are the most desirable for steganography because they have more space to hide the data. This field is expected to continually grow as fast development of computer graphics power and technology (Calpe,2006).

2.2 Image Steganography Methods

Image steganography has been widely studied by researchers. There are a variety of methods used in which information can be hidden in images. The following is the methods of image steganography:

2.2.1 Replacing Least Significant Bit

Least significant bit (LSB) insertion is a common and simple way to embedding information in a cover image (Johnson & Jajodia, 1998). For instance, a simple scheme proposed by Lee and Chen (2000), is to place the embedding data at the least significant bit (the 8^{th} bit) of each pixel in the cover image. The modified image is called stego-image. Manchanda,Dave and Singh(2007) found that altering LSB doesn't change the quality of image to human perception but it is sensitive to image processing attacks like compression, cropping etc. In addition, Nagham Hamid, Abid Yahya, R. Badlishah Ahmad & Osamah M. Al-Qershi(2012) discovered that changes of cover image using LSB techniques are very difficult to be recognized by the human eye because they are being too small.

2.2.2 Replacing Moderate Significant Bit

Chan and Chang(2001) showed the use of the moderate significant bits(the 4^{th} bit) of each pixel in the cover image to embed the secret message. This method improves sensitivity to modification, but it degrades the quality of stego-image.

2.2.3 Transformation Domain Techniques

It is another familiar data hiding techniques by use the transformation domain of digital media to hide information. Functions such as the discrete cosine transform (DCT) and the discrete wavelet transform (DWT) are widely applied in this technique. The DCT transforms a signal from an image representation into a frequency representation, by grouping the pixels into 8×8 pixel blocks and transforming the pixel blocks into 64 DCT coefficients each (Krenn, 2000). According to Nagham Hamid, Abid Yahya, R. Badlishah Ahmad & Osamah M. Al-Qershi(2012), wavelet transform clearly partitions the high-

frequency and low-frequency information on a pixel by pixel basis therefore wavelet is used in image steganography. Manchanda,Dave and Singh(2007) stated that the messages are hided in the significant areas of the cover image by using these methods, which makes them robust against image processing attacks like compression and cropping.

2.3 Steganalysis Techniques

The goal of steganalysis is to identify suspected information streams, determine existence of any hidden messages and recover the hidden message if possible (Si, 2004). A steganalyst is trying to determine the existence of a hidden message instead of knowing which bits carry what information (Zhang & Ping, 2006). There are several forms are taken to attacks and analysis on hidden information for example detecting, extracting and disabling or destroying hidden information (Chandramouli & Memon, 2006). Curran and Devitt(2008) indicated that steganalysis can be classified into two categories which are Passive Steganalysis and Active Steganalysis. Passive Steganalysis only involves detection while Active Steganalysis process is complete only after the hidden data is removed, destroyed or strategically altered to render it useless. Provos and Honeyman(2006) concluded that the main purpose of steganography is failed once its existence is revealed by steganalysis although the secret content is not exposed.

2.3.1 Steganalysis attacks

There are three types of Steganalysis attacks:

2.3.1.1 Visual or Aural attacks

They consist of striping away the significant parts of a digital content in order to facilitate a human’s visual inspection for anomalies (Wayner, 2002). The idea of Visual attacks is to remove all parts of the image covering the message so human eye can now distinguish whether there is a hidden message or still image content (Westfeld and Pfitzmann, 1999). A common test is to show the LSBs of an image.

2.3.1.2 Structural attacks

Sometimes the format of the digital file changes as hidden information is embedded so these changes lead to an easily detectable pattern in the structure of the file format (Westfeld and Pfitzmann, 1999). Identifying those characteristic structure changes can detect the presence of hidden file, for example in palette based steganography the palette of image is changed before embedding data to reduce the number of colors so that the adjacent pixel color difference should be very less as the result, this shows groups of pixels in a palette have the same color which is not the case in normal images (Bennett, 2004).

2.3.1.3 Statistical attacks

These types of attacks are more effective and successful as they reveal the smallest alterations in an images statistical behavior (Bhattacharyya & Sanyal, 2012). Statistical tests can reveal modified image by determining an image's statistical properties deviate from the norm (Provos & Honeyman, 2006). Digital pictures of natural scenes have distinct statistical behavior. Mercuri(2004) discovered that we can determine whether or not an image has been altered with proper statistical analysis, making forgeries mathematically detectable. Therefore, this Steganalysis method is to collect statistical evidences about the presence of hidden messages in images, and use them to verify the existence of hidden content on given images (Bishop, 2006).

2.3.1.4 Classification of attacks based on information available (Johnson, 2000)

- **Stego only attack**: Only stego object is available for analysis.
- **Known cover attack**: Both cover and stego are known.
- **Known message attack**: In some cases message is known. Analyzing the stego object pattern for this message embedded may help to attack similar systems.
- **Chosen stego attack**: Steganographic algorithm and stego object are known.
- **Chosen message attack**: Steganalyst implements many steganographic tools for a chosen message and analyses these stego objects with the one which is to be analyzed and try to find the algorithm used in these process.
- **Known stego attack**: Cover, object and the steganographic tool used are known.

	Stego object	Original cover object	Hidden message	Stego algorithm or tool
Stego only attack	x			
Known cover attack	x	x		
Known message attack	x		x	
Chosen stego attack	x			x
Chosen message attack	x			
Known stego attack	x	x		x

Table 2.1 Summary of attacks based on information available

2.3.2 Steganalytic Methods

Steganalytic Methods	**Description**	**Targeted Steganographic Techniques**
RS (Regular and Singular groups) steganalysis	Sensitivity of dual statistics based on spatial correlation of pixels to LSB randomization due to steganographic embedding is used in analysis.	Various LSB modification techniques
PoV(Pairs of Value)-based Chi-square test	Chi-square test checks whether the occurrence of each pair of values tends to become equal, indicating some data is embedded.	Steganography based on swapping pairs of values of pixel gray levels, colors, or DCT coefficients
Palette checking	Peculiarity in palette ordering is a clear sign of systematic modification.	Steganography in palette images
RQP (Raw Quick Pairs) method	Method based on analyzing the increased number of close-color pairs caused by embedding.	LSB embedding in true-color images
Check JPEG compatibility	Method detects unusual departure from the JPEG signature inherent in images initially stored in JPEG format.	Space-domain steganography using images initially

		stored in the JPEG format
Histogram analysis	Method reveals discreteness or periodicity in particular coefficients due to quantization-related modification.	QIM (Quantization Index Modulation) or other quantizationrelated embedding methods
Universal blind detection	Statistical quantities constructed using high-order statistics, and a detection model established with the threshold obtained in a training process.	Various steganographic techniques
Peak Signal-to-Noise Ratio (PSNR) analysis	Calculation of Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) values to identify concentration of the noise inside the stegoimage.	LSB modification technique.

Table 2.2 Comparison of Steganalytic methods (Generate from Wang and Wang, 2004)

2.4 Conclusion

Peak Signal-to-Noise Ratio (PSNR) steganalysis is chose as the method used in my program because it is use for detect various LSB modification techniques. LSB is the method that most commonly used in steganography to hide message. So, it is necessary to develop a program to detect it.

Chapter 3:

Methodology

3.1 Introduction

Methodology is an organized, documented set of procedures and guidelines for problem solving with components like techniques, tools, methods, and tasks. Methodology includes a diagramming process for documenting the outcomes of the procedures, approach for carrying out the procedure, and determining the quality of results of the procedures. The methodology may include publication research, questionnaires, interviews, surveys and other research techniques, and may include present and previous information. It describes the methods to be used. Research design, the population to be studied, and the research instruments or tools to be used are discussed in the methodology. In short, methodology is the way of how to conduct research.

3.2 Research Approach

Research approach is the method chosen for development of the application which includes describing the process used to develop software or system product. It describes the activities performed and how the development phases follow each other to ensure the success in the process of system development. Agile modeling is chosen to ensure the application development process runs smoothly.

3.2.1 Agile Modeling

Agile modeling is a framework or process that describes the activities performed at each stage of a software and system development project. Agile modeling methodology divides the software development process into several phases to make the whole process become more organize and easier to achieve, control, and manage thus help out software developer to develop an organized system. The main purpose of Agile modeling is to produce a high quality system that meets or exceeds customer requirements or expectations, complete works within time and cost estimates, works effectively and efficiently by utilizes current and planned Information Technology infrastructure, and cost-effective to maintain and enhance. Agile modeling ensuring that all functions, user requirements and agency strategic goals and objectives are met. Each phase of agile modeling continues and refines what is done in the previous phase. Commonly known development phases in agile modeling are planning, analysis, design, implementation, verification and maintenance.

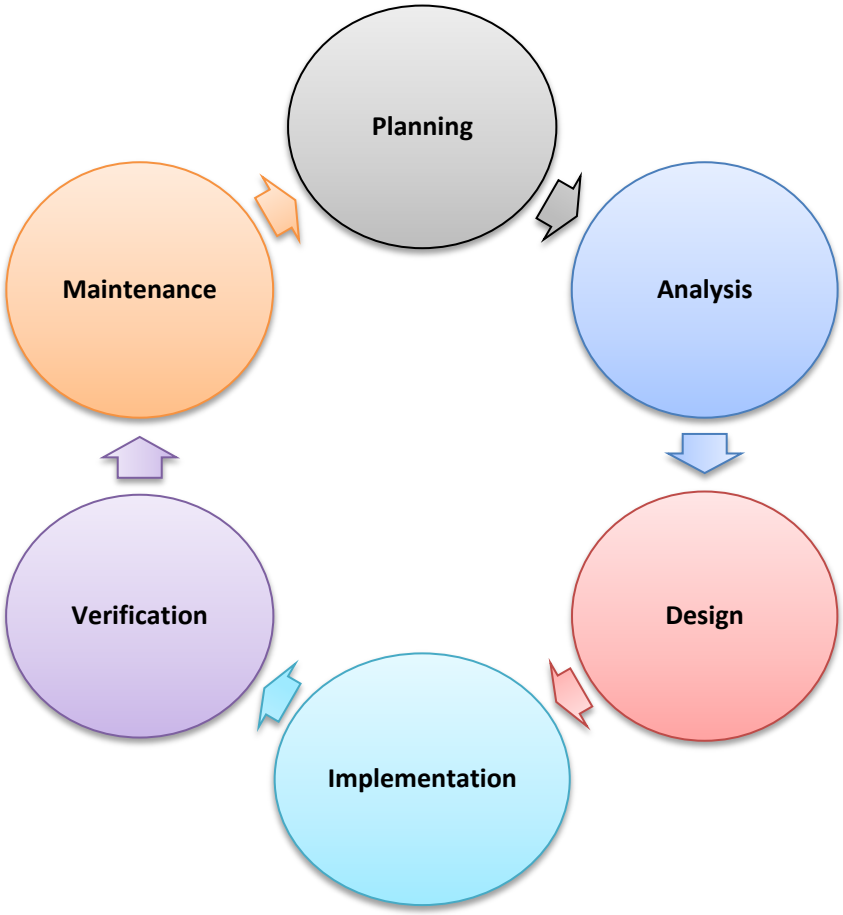

Figure 3.1 Phases in Agile Modeling

Agile modeling is chosen because it brings benefits to system development. Firstly, agile modeling can avoid unexpected high cost and low expected benefits. It allows the progress to be reviewed at the conclusion of each phase. Furthermore, agile modeling has detail steps and processes and well defined user input. For agile modeling, deadline date is more adhered because the completion time is fixed. The other advantages of agile modeling include evaluate costs and completion targets, development and design standards are good, relevant software and system documentation, maximize productivity, and improve quality of systems. Besides that, the system development process will become easy to monitor, manage and control as well for maintenance. Compare with waterfall methodology, agile modeling is more flexible and it allows backward process to change the decisions and works that had made in previous phase after each phase is completed and it is welcome for changing requirements, even late in development.

3.3 Implementation of the Project

Development phases in agile modeling are planning, analysis, design, implementation, verification and maintenance. The details and the description of each phase are important to developer to understand activities of each phase and the progress of the project.

3.3.1 Planning Phase

It is the process of understanding why the system should be built, determines the goals and objectives of the project, and defining its requirements. It also includes feasibility study from several different perspectives, technical, economic, and organization feasibility aspects. A project management plan and other planning documents are developed. Provide the basis of requiring the resources needed to achieve solution. The risks and various project-planning approaches are defined. The existing system is evaluated and deficiencies and weaknesses are identified during this phase. The examples of existing systems to be studied are StegDetect, Stegspy and Xsteg. Gantt chart is produced during this phase. All the tasks are planned and arranged. The timeline of the project is shown in Appendix A.

3.3.2 Analysis Phase

Information needs and requirements of the end users, project goals, organizational environment, and any system presently being used are analyzed and also develop the functional requirements of a system that can meet the needs of the users. Besides that, the requirements are recorded in a document. The requirements documentation should be referred to throughout the rest of the system development process to ensure the developing project meets and fulfill the user needs and requirements. Problems are identified and suggestions are recommended for improving the system functioning. For example, find out problems exist and attempt to fix the system. During this phases, the problems are found in existing systems are they do not have GUI and not suitable for new users to use as they does not have enough technical knowledge to operate it. Suitable technique of steganalysis and tools is identified during this phase. The PSNR technique is chosen to detect LSB modification in image files and tools suitable for develop a new application is MATLAB.

3.3.3 Design Phase

During this phase, all detail functional requirements are translated into preliminary and complete designs. Decisions are made to address how the system will meet functional requirements. A general system design emphasizes the functional features of the system. After that a final or detailed system design is produced which specifying all the technical detail needed to develop the system. There are several techniques used for describing the system design of the system for example flowchart and data flow diagram (DFD). The design describes desired features and operations in detail, including screen layouts, business rules, process diagrams, pseudocode and other documentation. Prototype interface design and use case diagram are created during this phase. Use case diagram describes the relationship among the application, user and developer. GUI is implemented to the application so that new users can operate it more easily.

3.3.3.1 Prototype Interface Design

Figure 3.2 Prototype interface design

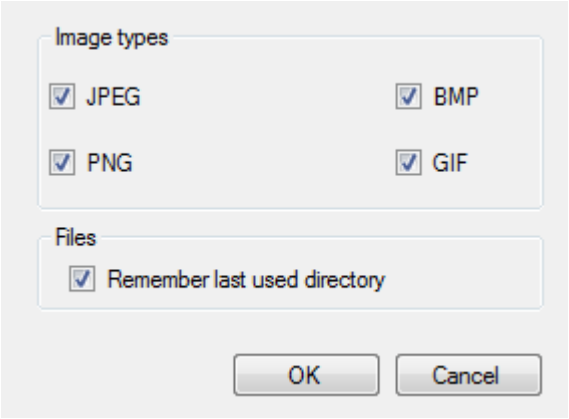

Figure 3.3 Prototype option interface

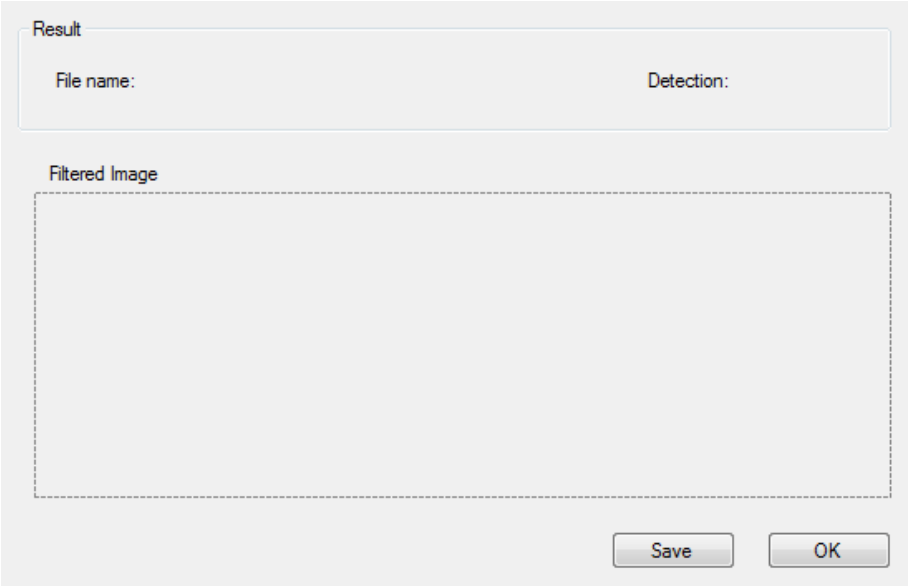

Figure 3.4 Prototype result interface

The “open file” button is used to choose an image file to test. The test button is used to test the chosen image. Click the test button the result interface will come out and show filtered image, file name and detection result. In the result interface, users can click save button to save the results. The clear button is used to clear all the information and chosen image. Exit button is used to exit the software. The input file information of chosen file like directory, file name, file size, and pixels is show in the interface. The image will show in the interface also. Click the “option” function, the option interface will pop out and the users can choose desirable option for image type or remember last used directory.

3.3.3.2 Flow Chart

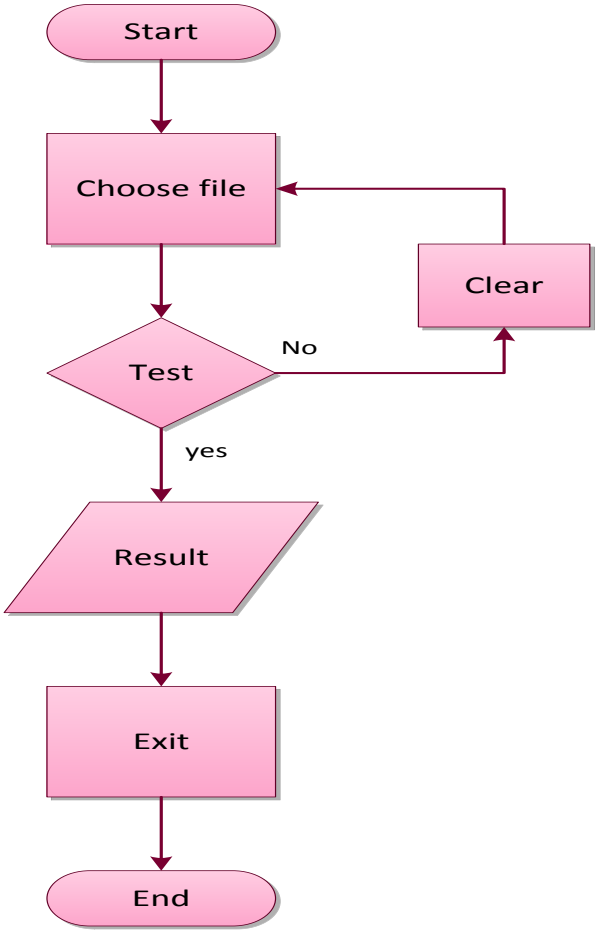

Figure 3.5 Flowchart of system

First, user start with execute the software. Then, choose an original image file (JPEG/JPG, PNG format) and suspected image to test. After image files are chose, choose the test function for testing and analyzing the image file. If the user chooses a wrong file, user can clear the chosen image and choose a new image file. After test, result will come out and show on the interface. Finally user can exit the software if not using anymore.